基于核心作者研究兴趣相似性网络的社群隶属研究*

——以国内情报学领域为例

■ 徐健[1] 毛进[1] 叶光辉[2] 巴志超[1] 李纲[1]

[1] 武汉大学信息资源研究中心 武汉 430072 [2] 华中师范大学信息管理学院 武汉 430079

摘要：［目的/意义］构建情报学领域核心作者研究兴趣相似性网络，研究该网络的社群结构与社群隶属问题。［方法/过程］在 CSSCI 数据库以中图分类号为检索条件，下载该学科 1998－2015 年所有的论文数据，通过普莱斯定律识别核心作者。将作者研究兴趣用词袋模型表示，并计算作者间研究兴趣的余弦相似性，进而构建核心作者研究兴趣相似性网络。在此基础上进行社群划分并识别各社群研究主题，计算作者对各社群的隶属度与模糊熵。［结果/结论］研究发现，当前我国情报学研究可分为信息组织与检索、文献计量与科学评价、竞争情报与知识管理、情报学学科整体研究 4 个领域，大部分作者研究并不局限于单个领域，竞争情报和文献计量领域界限明显，较少有作者将彼此作为次要隶属社群。

关键词：研究兴趣 网络 情报学 社群 隶属度 相似性

分类号：G250

DOI:10.13266/j.issn.0252－3116.2018.12.008

1 引言

探查学科内作者社群结构，可以把握其领域分布及未来发展方向，对科研活动的管理、组织与协调具有重要意义。社群发现方法就是从复杂网络中发现具有模块结构特性的群体，进而发现社群随时间变化的趋势、规律与动因[1]。该方法在学科结构分析中具有广泛的应用。直观来看，核心作者的研究兴趣对学科发展方向具有重要影响，因此从其研究兴趣角度揭示学科结构、研究热点相比其他方法来讲更具有一定优势。为此，笔者将学科的社群结构定义为由研究兴趣相似的若干作者组成的群体。

传统的社群发现方法均遵循"每一个节点都唯一归属于某个社群"的假设，而在现实社会网络中，人们往往同时归属于不同的社群，这些人往往又是信息传递、社会交往中的关键节点[2]。同样的问题也存在于学术社群研究中，笔者以情报学为例，通过识别该领域核心作者，构建其研究兴趣相似性网络，在此基础上开展社群结构与社群隶属研究。本文重点关注以下问题：情报学领域核心作者研究兴趣相似性网络有哪些特征？各个社群研究主题是什么？核心作者对不同社群的隶属度分布模式是怎么样的？不同社群之间是否有关联，关联模式如何？通过对上述问题的回答，以期为相关研究提供借鉴与参考。

2 相关研究

2.1 情报学学科结构分析

关于情报学领域结构的分析存在两种研究思路：定性研究和定量研究，前者依靠作者对学科整体情况的宏观把握，重点关注范式归纳[3]、基本原理总结[4]、阶段划分[5]等问题，这类研究属于思辨性研究。后者多利用文献计量学、社会网络分析方法深入挖掘文献数据中的研究主题、社群结构，并以可视化的方式展现结果。例如，张斌等[6]利用共词分析方法对我国图情领域认知结构进行分析与识别，认为国内图书情报学主要存在 4 个分支：数字信息的组织与管理、图书馆

* 本文系国家自然科学基金项目"科研团队动态演化规律研究"（项目编号：71273196）、"基于排名分析－特征识别的应急决策专家发现"（项目编号：71740012）和"突发公共卫生事件社交媒体信息主题演化与影响力建模"（项目编号：71603189）研究成果之一。

作者简介：徐健（ORCID：0000－0002－0230－5137），博士，E-mail：xukeywhu@163.com；毛进（ORCID：0000－0001－9572－6709），博士后；叶光辉（ORCID：0000－0002－4162－2282），讲师；巴志超（ORCID：0000－0001－5626－5604），博士研究生；李纲（ORCID：0000－0001－5573－6400），主任，教授，博士生导师。

收稿日期：2017－09－26 修回日期：2018－03－09 本文起止页码：57－64 本文责任编辑：徐健

学、信息计量与评价和组织知识管理。马费成等[7]利用 ACA 分析我国 1994 – 2005 年间情报学核心作者共被引情况，得出国内情报学研究领域的 5 个领域：早期研究者、情报学理论、情报检索、图书馆学研究、文献资源建设。邱均平[8]对我国改革开放 30 年来情报学研究论文的作者进行分析，将改革开放以来情报学研究内容划分为竞争情报、信息系统、文献计量等 8 个领域。上述研究多利用共词网络、合著网络、共被引网络等从学科整体层面识别社群主题，较少有从研究兴趣角度出发开展领域结构识别的研究，更少有研究深入微观层面探讨学者与社群间隶属关系。

2.2 研究兴趣相关研究

兴趣作为一种心理活动，对人们认知倾向和实践活动具有重要的影响。在科学研究过程中，分析某领域核心作者研究兴趣有助于把握该领域现状与未来发展方向。当前关于研究兴趣有关的研究主要涉及研究兴趣表示模型及相似性计算、研究兴趣现象揭示两方面。

在研究兴趣表示模型及相似性计算方面，M. Steyvers 等[9]对原始主题模型进行修改，提出作者主题模型（Author Topic）用于揭示作者研究兴趣，并探索该模型在主题演变趋势分析、作者主题关联分析、作者异常论文检测等方面的应用；李树青等[10]以向量空间模型为兴趣基本表达结构，并采用时间片震荡算法发现作者主要的研究兴趣特征，开展便携式个性化服务研究；刘萍等[11]通过向量空间模型计算关键词间关联，并利用 P-Rank 算法计算两个作者关键词网络的结构相似度；巴志超等[12]引入 word2vec 模型对作者关键词矩阵进行语义建模，计算两个作者研究兴趣矩阵的 JS 距离作为其兴趣的相似性，并通过聚类算法将研究兴趣相似的作者聚类形成类团。

在研究兴趣现象揭示方面，李纲等[13]使用词袋模型表示作者研究兴趣，通过合著作者研究兴趣之间的相似性，分析了不同学科、不同产量的合著作者研究兴趣相似性频率分布模式。研究发现，不同学科的作者在寻找合作伙伴时，在研究兴趣相似程度的倾向性的分布规律上具有一致性，即随相似度升高，先升高后减小。而高产作者更倾向于寻找研究兴趣更相似的作者进行合著。关鹏等[14]结合生命周期理论与作者主题模型对 CNKI 数据库内锂电池领域作者研究兴趣演化方式进行了分析。研究发现，核心作者的研究兴趣演化趋势与相应的主题演化趋势一致时，能够引领该研究主题的发展。

综上所述，当前关于研究兴趣的研究主要涉及研究兴趣表示模型及相似性计算、研究兴趣现象揭示两方面。研究兴趣表示模型方面主要有主题模型、词袋模型、关键词网络、关键词语义矩阵 4 种，研究兴趣相似性计算方法与具体表示模型有关，主要有余弦相似度、JACCARD、P-rank、JS 距离等。在研究兴趣现象揭示上，当前研究主要有合著与研究兴趣相似性的关系、研究兴趣的发展演化等。笔者以情报学为例，通过识别该领域核心作者生成其研究兴趣相似性网络，对该网络特征进行分析，在此基础上进行社群结构分析，计算各作者对各社群的隶属度，分析各社群之间的关联。

3 研究方法

3.1 研究思路与研究过程

从前文可知，社群分析中将节点唯一的划分到一个社群中有其不合理之处。同样在科研领域，研究人员的研究兴趣也很少限定于某个领域，而是按照一定的程度同时隶属于多个学术社群。笔者借鉴文献[1]中的计算思路，首先对网络进行社群划分，然后计算各个作者对各个社群的隶属程度。具体研究过程分为数据获取与处理、研究兴趣相似性网络构建、社群识别与隶属度计算三个步骤，如图 1 所示：

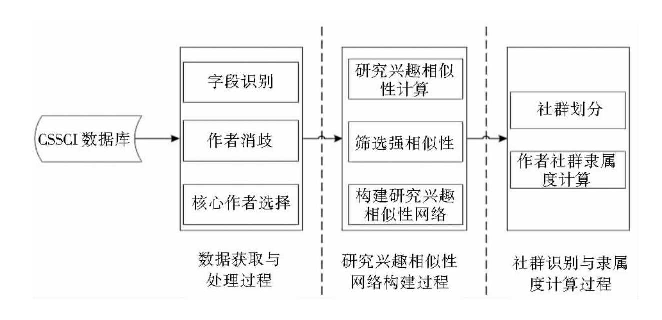


图 1 本文研究过程

3.2 数据获取与处理

数据获取与处理主要包括以下步骤：数据下载、字段识别、作者消歧、核心作者选择。数据源为南京大学社会科学引文索引数据库，检索条件为“中图分类号 = G35”。使用 Java 程序从该题录数据中抽取出题目、作者、作者单位、期刊名称、关键词、期刊名等字段并将之存入到数据库中。为保证数据分析的准确性，作者消歧过程采用系统消歧与人工校对相结合的方式。具体步骤如下：首先将“姓名 + 作者一级单位”的作者判定为同一位作者，其次人工逐个分析姓名相同而单位不同的情况，根据实际情况对这些作者进行合并。为保证作者间兴趣相似性计算过程中的作者关键词数目的

数量,笔者拟以学科核心作者作为研究对象构建网络。根据普莱斯定律[15],将发文量等于以及大于 N 篇的作者为某领域的核心作者,N 的计算公式见公式(1):

$$N = 0.749 \times (\eta_{max})^{1/2} \qquad 公式(1)$$

其中 η_{max} 表示该领域发文量最大的作者文章数目,在本数据集中,发文量最高的为武汉大学邱均平教授,其发文 198 篇,此时 N 取值为 10.384。由此,本数据集中发文量大于等于 11 的作者均为核心作者,共有 220 位。

3.3 研究兴趣相似性网络构建

研究兴趣相似性网络以作者为网络节点,以作者间的研究兴趣相似度为边权重,其构建过程主要包括以下步骤:研究兴趣相似性计算、强相似性筛选、研究兴趣相似性网络构建。

本文使用的研究兴趣相似性计算方法可参考文献[16]中利用合著双方作者私有关键词集的余弦相似度算法,该计算方法可以有效去除合著对研究兴趣相似性计算的影响。在计算相似度之前,要先计算各关键词的逆文档频率(具体计算过程见公式 2)。在公式(2)中,k_i 是一个关键词,N_{doc} 为文档总数,$df(k_i)$ 为关键词在多少文档中出现。

$$idf(k_i) = \ln\left(\frac{0.5 + N_{doc}}{df(k_i)}\right) \div \ln(1 + N_{doc}) \qquad 公式(2)$$

假设有 A、B 两位作者,则 P(A)、P(B)分别代表两位作者发表文献的集合。若 A 与 B 曾发生过合著关系,则分别从 P(A)、P(B)中剔除两者合著文献集合,生成 A、B 两者私有文献集合 P(A)、P(B)。此时,KA 是文档集 P(A)形成的关键词列表,$N_A(k_i)$ 和 $N_B(k_i)$ 表示关键词 k_i 在 K_A 与 K_B 中出现的次数,最终的研究兴趣相似性计算见公式(3)。通过两两计算核心作者的研究兴趣相似性,可以得到作者相似性矩阵。

$$sim(A,B) = \frac{\sum_{k_i \in K} W_A(k_i) \times W_B(k_i) \times idf^2(k_i)}{\sqrt{\sum_{k_i \in K_A}(W_A(k_i) \times idf(k_i))^2} \times \sqrt{\sum_{k_i \in K_B}(W_B(k_i) \times idf(k_i))^2}} \qquad 公式(3)$$

直观看来,两位作者发文中共同使用过一个单词,其相似性也不为零。使得这样生成的网络必定很密集,此时需要进行剪枝操作,剔除权重较低的边。此时,需要根据一定策略对网络中作者之间的边进行筛选。为便于后续社群分析,将研究兴趣相似性矩阵进行剪枝,并将该网络保存为 Gephi 分析的数据格式。

3.4 社群识别与隶属度计算

在研究兴趣相似性网络构建完成后,即可使用网络划分方法对该学科研究领域进行划分,进而计算作者对社群的隶属度。该过程主要包括以下步骤:社群结构识别、作者－社群隶属度计算、基于社群隶属度的模糊熵计算等。本文社群划分方法使用 Louvain 算法[17],该方法可以快速处理具有数亿节点的网络,利用 Gephi 软件将网络中核心作者划分到不同社群中。学者的研究兴趣在不同阶段会发生变化,同时不同领域之间存在关联与交叉的部分,这使得很多作者的研究兴趣跨越两个甚至更多领域[7]。因而,笔者引入模糊数学中"隶属度"的概念来表示作者研究兴趣在不同社群中的分布,以解决作者唯一地归属某社群的问题。此处定义作者对社群的隶属度为作者对各个社群相似度的归一化结果,具体计算方法见公式(4):

$$membership(A_i, C_j) = \frac{Similarity(A_i, C_j)}{\sum_{k=1}^{4} Similarity(A_i, C_k)} \qquad 公式(4)$$

学者对不同社群的隶属度可以看作其研究兴趣的离散分布,因此可以使用该分布的熵表示该作者研究兴趣的分散程度。熵的概念最初来自物理学领域,在信息领域经常用于表示信息量度量,用于表示信息源的不确定性,基于社群隶属度的模糊熵计算过程见公式(5):

$$Entropy(A_i) = -\sum_{j=1}^{n} membership(A_i, C_j) \times \log_2(membership(A_i, C_j)) \qquad 公式(5)$$

在公式(5)中,n 表示社群数目,在本研究中,n 取值为 4。当作者的研究兴趣均衡的分布在各个社群中时,此时其研究兴趣比较分散,很难确定作者的主要研究兴趣,此时,熵值较大。

4 实验结果

4.1 原始数据描述

截至笔者下载该数据集时,该数据库中 2016 年论文数据还未被完全收录,故从中下载 1998－2015 年间的全部 14 530 篇文章题录数据。经统计数据集共来自 257 种期刊,其中,《情报杂志》《情报科学》《情报理论与实践》《图书情报工作》《情报学报》5 家期刊位列载文量前 5 位。从第一作者发文单位来看,武汉大学、南京大学、中信所、吉林大学、北京大学排名前 5 位。

4.2 研究兴趣相似性网络特征分析

核心作者研究兴趣相似性网络以作者为网络节点,以作者间的研究兴趣相似度为边权重。据统计该

网络共包含 220 个节点,18 525 条边,网络密度为 0.769。这说明该网络中连接线较密集,社群识别难度较大,需要对该网络进行剪枝。作者研究兴趣相似度是连续型变量,其值介于 0 到 1 之间。将该连续性数值进行离散化操作,划分为 100 组,组距为 0.01,组数为 20。研究兴趣相似性最大的两位作者是赵筱媛与陈峰,二人研究方向均是产业竞争情报。对情报学 220 位核心作者形成的研究兴趣相似性网络的边权重进行幂律分布拟合,如图 2 所示:

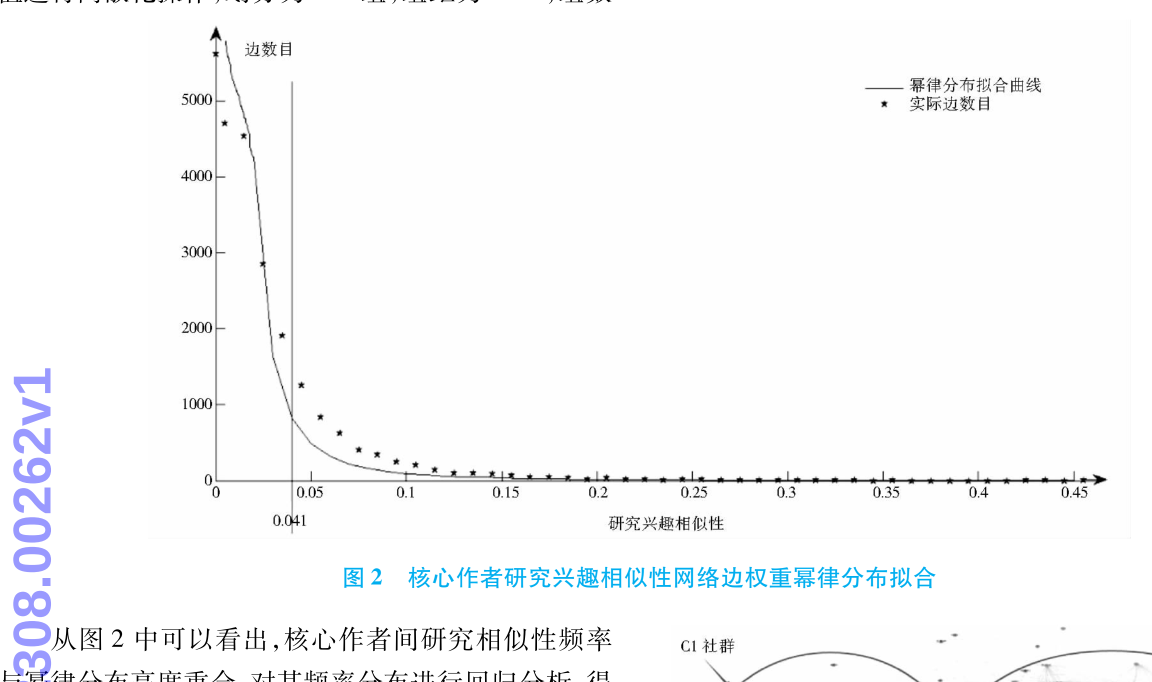


图 2 核心作者研究兴趣相似性网络边权重幂律分布拟合

从图 2 中可以看出,核心作者间研究相似性频率与幂律分布高度重合,对其频率分布进行回归分析,得到相似性(记为 x)的分布密度函数为公式(6)。此时对研究兴趣相似性网络进行剪枝处理,根据"二八定律",需要剔除权重较低的 80% 的边。在实际操作中,本研究定义 0.041 为相似性强弱的分界点,仅保留大于权重大于 0.041 的边进行分析。剪枝后的节点共包含 220 个节点,4 563 条边,此时网络密度 0.189。

$$f(x) = 0.4018 \times x^{-2.3691} \qquad \text{公式(6)}$$

4.3 各个社群研究主题分析

使用 Gephi 软件对剪枝后的网络进行模块度计算,设置 Resolution 为默认值 1.0,得到情报学研究兴趣相似性网络可以分为 4 个社群,模块度为 0.347。相关研究表明,模块度值在 0.3 - 0.7 时网络中会出现较强的社群结构[18]。社群识别结果见图 3。

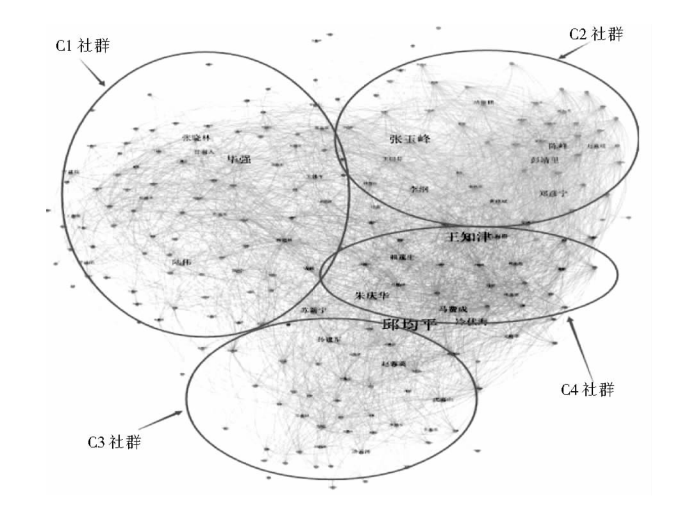


图 3 社群划分可视化结果

得到的情报学领域所有作者被聚集成 4 个社群,分别记为 C1、C2、C3、C4,网络中节点大小设置为作者发文量。从图 3 中可以看出,这 4 个社群之间界限并不分明,发文量较大的作者一般位于社群交界处,这些作者研究兴趣较为广泛,是各个社群沟通交流的关键节点。由于本文是将研究兴趣相似的作者群体进行聚类的,因此识别出的社群应该拥有各自的研究主题,而不同社群间研究主题迥异。笔者为进一步探究各社群研究主题,分别绘制各社群作者所著文献高频关键词共现网络,分别如图 4 至图 7 所示:

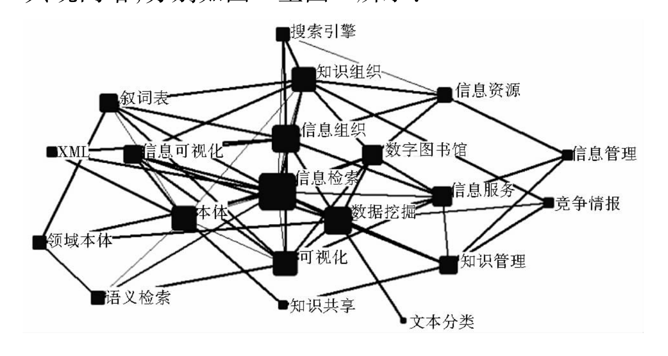


图 4 社群 C1 高频关键词共现网络

C1 社群作者主要研究包括信息组织、信息检索、本体、数据挖掘、可视化、语义检索在内的信息处理技术，该领域作者有 86 位，约占核心作者总数的 39%，是情报学领域的主流研究群体。该领域技术性较强，与计算机科学、信息系统等领域研究交叉较多，是情报学领域最具特色的领域之一。这些方法与技术在当今大数据时代背景下仍然具有较强的解决问题的能力，未来该方向的研究可更多结合人工智能技术，不断实现技术上的创新。

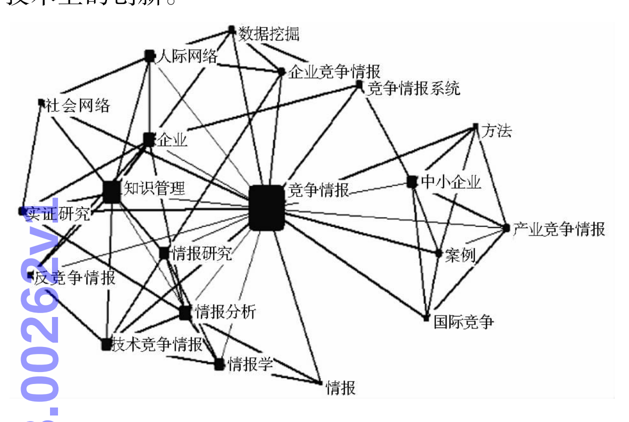


图 5　社群 C2 高频关键词共现网络

C2 社群作者主要研究竞争情报，该领域研究起源于 20 世纪 80 年代，在情报学中占据重要地位。竞争环境、竞争对手和竞争策略是竞争情报的三大要素。当前该领域研究主要集中在竞争情报与知识管理的关系、反竞争情报、竞争情报技术与系统、产业(企业)竞争情报等方面。

C3 社群作者主要研究文献计量与科学评价问题，该领域发展历经书目统计、文献计量、信息计量、网络计量等阶段[8]。文献计量是情报学定量研究的重要内容，具有悠久的发展历史，相继出现了文献计量学、科学计量学、信息计量学、网络计量学、知识计量学，可以统称为“五计学”。目前，替代计量学(Altmetrics)是国际计量学领域的热点问题，目前得到了国内学界的普遍关注。

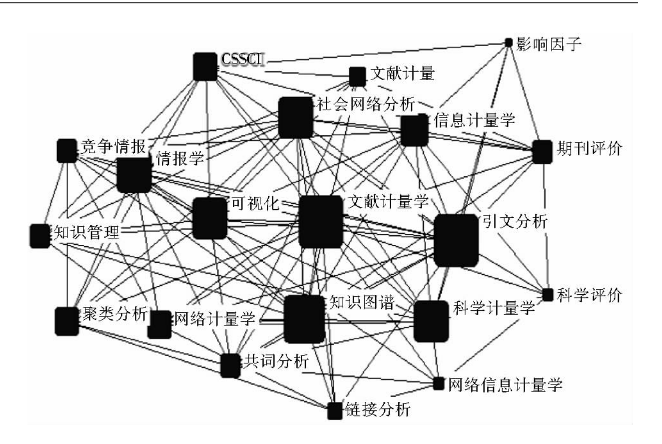


图 6　社群 C3 高频关键词共现网络

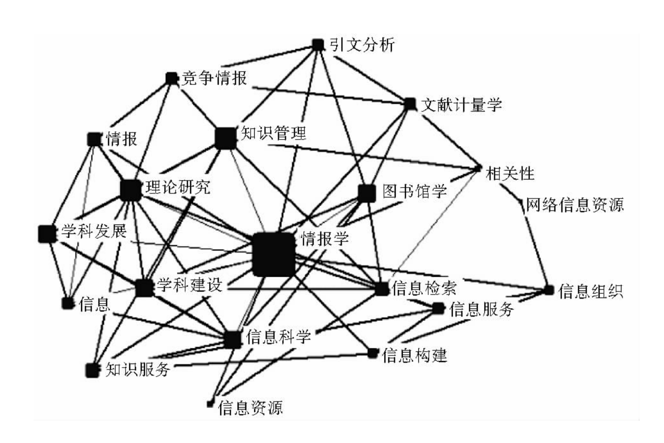


图 7　社群 C4 高频关键词共现网络

C4 社群作者以情报学学科整体为研究对象，重点研究情报学理论、学科边界与发展，发现本学科与信息科学、图书馆学等相关学科的关联。同时，该社群的关键词包含了文献计量、信息检索、竞争情报在内其他社群的高频关键词。这说明，竞争情报、文献计量、信息检索等作为情报学领域的子领域，其研究与学科的整体研究紧密相关。

表 1 社群识别基础描述性统计结果，包括各社群网络节点数目、强相似性边数目、密度、代表作者。

表 1　社群识别基础描述性统计结果

社群编号	节点数目	社群内强相似数目	社群密度	代表作者(发文量:篇)
C1	86	703	0.192	毕强(57)陆伟(40)张晓林(40)甘利人(40)周宁(35)贾君枝(35)
C2	46	744	0.719	张玉峰(88)李纲(64)郑彦宁(59)王曰芬(44)彭靖里(42)靖继鹏(40)
C3	54	512	0.358	邱均平(198)朱庆华(74)冷伏海(69)苏新宁(57)孙建军(53)赵蓉英(49)
C4	34	439	0.783	王知津(129)赖茂生(52)马费成(45)马海群(43)成颖(33)梁战平(30)
整体	220	4 563	0.189	—

从表 1 中可以看出，各社群内部研究兴趣相似性网络的密度均高于平均密度，其中 C1 社群网络密度较低，说明该社群研究范围比较广泛，与其他学科有较多交叉，存在进一步划分的可能。从上文中可以看出，笔者识别出 4 个社群的研究并非完全隔绝，而是互相渗透与借鉴。

4.4 核心作者社群隶属度分析

由前文可知，作者与社群并非一一对应的，一个作者在其研究的不同阶段其研究兴趣可能发生变化。因此，笔者引入隶属度的概念，计算作者与社群及与各个社群研究的归属关系，并据此计算反映作者兴趣分散程度的模糊熵。部分作者对各个社群的隶属情况如表 2 所示：

表 2 作者－社群隶属度

姓名	C1 隶属度	C2 隶属度	C3 隶属度	C4 隶属度	基于社群隶属度的模糊熵
邱均平	0.12	0.15	0.52	0.21	1.90
王知津	0.12	0.25	0.14	0.48	1.93
张玉峰	0.24	0.49	0.13	0.14	1.93
…	…	…	…	…	…

表 2 中每行分别对应作者对不同社群的隶属度以及模糊熵。220 位核心作者隶属度最高的社群均为前文社群划分结果完全一致。基于社群隶属度的模糊熵反映了作者研究兴趣的分散程度。在本研究的所有作者中，大连理工大学的王贤文模糊熵，最小为 1.39，对 C3 社群的隶属度高达 80%，其研究主要集中在文献计量与科学评价领域。安徽大学储节旺模糊熵最大，为 2.05，对所属社群 C1 的隶属度仅为 36.4%，其研究广泛涉猎竞争情报、文献计量等领域。此外，笔者绘制了 220 核心作者模糊熵的频次分布折线，如图 8 所示：

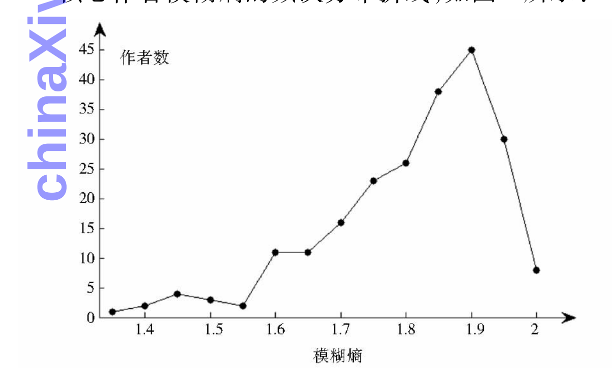


图 8 模糊熵核心作者人数分布

从图 8 中可以看出，情报学领域核心作者兴趣隶属模糊熵集中在 1.6－2 之间，这说明当前大部分研究者研究范围涉及多个领域。这种情况的产生可能有两个原因：①学者研究兴趣会随科学研究的进展而发生变化；②现阶段越来越多研究需要多个学科协同开展，因此作者需要同时涉猎多领域的知识。对核心作者按照模糊熵大小进行分组，得到各组作者平均发文量见图 9。

从图 9 中可以看出，作者平均发文量与模糊熵呈

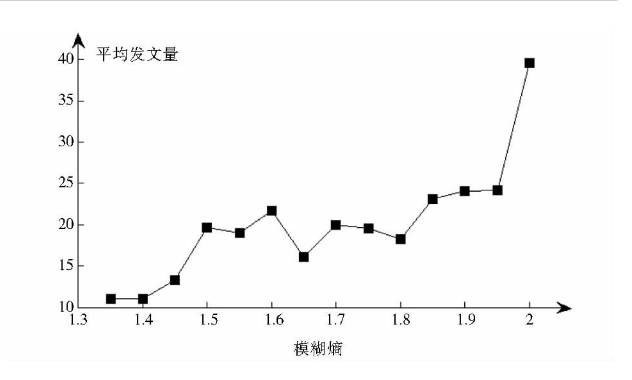


图 9 基于模糊熵分组的作者平均发文量

正相关关系。这说明，高产作者的研究兴趣一般比较广泛，研究方向较为分散。各社群成员对各社群的平均隶属度如图 10 所示：

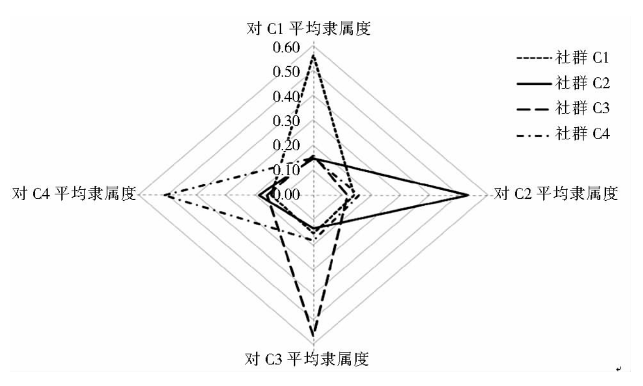


图 10 各社群作者对各社群平均隶属度雷达示意

从图 10 中可以看出，各社群作者对各自所属社群的隶属度平均值均位于 0.5－0.6 之间，对其他社群隶属度的均值位于 0.1－0.2 之间。各社群作者模糊熵均值位于 1.8－1.9 之间，这说明各社群作者研究兴趣分散程度较一致，差异不大。

4.5 作者次要隶属社群分析

从前文可知，作者按照一定隶属度归属于某个社群，由此可以对单个作者重叠的多个社群进行排序。在此，定义作者的次要隶属社群为作者隶属度第二大的社群，本部分重点关注不同社群作者的次要隶属社群，以此发现社群之间的关联关系。各社群作者的次要隶属社群分布如表 3 所示：

表 3 各社群次要隶属社群比例

社群编号	该社群用户第二兴趣
C1	C4(41%)C3(33%)C2(26%)
C2	C4(59%)C1(26%)C3(15%)
C3	C4(46%)C1(41%)C2(13%)
C4	C3(39%)C1(32%)C2(29%)

从表 3 中可以看出，C1-C3 社群的作者均将 C4 社群研究主题作为自己的次要隶属社群。从前文可知，

C1 社群作者以信息组织与检索为主,C2 社群作者以竞争情报为主,C3 社群以文献计量与科学评价为主。在情报学中,这三个研究领域的研究方法与范式迥异,边界比较明显,而 C4 领域着眼于情报学层面,其研究视角和层次均要高于其他社群,同时该社群与其他三个社群又有紧密联系。同时,笔者还注意到 C2 社群和 C3 社群的作者很少将彼此作为自己的次要隶属社群。竞争情报领域作者研究注重理论与实践结合,具有较强的应用价值,而文献计量与科学评价领域作者更注重研究问题的学术性,偏重于数据分析方法与实验结果,其研究更学术化一些。上述原因导致这两个社群之间关联度较低,交叉领域较少,这与图 5 中社群可视化结果 C2 与 C3 社群相互间距离较远的情形比较吻合。

5 结论

关于重叠社群的研究受到学界的广泛关注,笔者通过构建情报学领域研究兴趣相似性网络,与社群发现算法发现当前情报学研究的 4 大领域,利用各社群的高频词共现图谱得到各自研究主题。之后计算各作者对各社群的隶属度及模糊熵,将作者依据模糊熵分组,统计各组人数及人均发文量。最后,统计各社群作者的次要隶属社群分布。研究发现:①情报学核心作者研究兴趣相似性网络边权重符合幂律分布;②当前我国情报学研究可分为信息组织与检索、文献计量与科学评价、竞争情报与知识管理、情报学学科整体研究 4 个领域;③领域作者的研究兴趣多数并不局限于一个领域,而是分布于多个研究领域,同时,模糊熵越大的作者,其发文量也更多;④C1、C2、C3 社群的作者均将 C4 社群作为其次要隶属度矩阵,说明该社群与其他社群研究较为密切,存在较多交叉领域;⑤C2 与 C3 的作者很少将彼此作为次要隶属社群,说明这两个领域的作者研究范式差异较大,相互间交流较少。

需要注意的是,本文的数据来源受 CSSCI 收录数据的限制、作者消歧和选择方法还存在一些问题。未来将朝着以下几个方向开展研究:①现有作者研究兴趣相似性计算方法有待改进,需要考虑关键词之间的语义关系,对关键词多词一义的情况进行合并;②将时间因素考虑进去,分析研究兴趣相似性网络动态演化过程;③探索更为准确的社群识别方法,分析多个社群间信息交流与合作关系,扩展研究兴趣相似性网络应用范围;④从引文、合著等角度探究各社群间的交流与关联。

参考文献:

[1] 胡云,王崇骏,谢俊元,等. 社群演化的稳健迁移估计及演化离群点检测[J]. 软件学报,2013,24(11):2710－2720.
[2] 潘磊. 若干社区发现算法研究[D]. 南京:南京大学, 2013.
[3] 马费成. 情报学发展的历史回顾及前沿课题[J]. 图书情报知识, 2013(2):4－12.
[4] 马费成. 论情报学的基本原理及理论体系构建[J]. 情报学报, 2007,26(1):3－13.
[5] 马费成,宋恩梅. 我国情报学研究的历史回顾(Ⅰ)[J]. 情报学报, 2005,24(4):515－523.
[6] 张斌,贾茜. 我国图书情报学的认知结构及其演化[J]. 中国图书馆学报, 2014,40(4):30－47.
[7] 马费成,宋恩梅. 我国情报学研究分析:以 ACA 为方法[J]. 情报学报, 2006, 25(3):259－268.
[8] 邱均平,杨思洛,周春雷. 改革开放 30 年来我国情报学研究论文的作者分析[J]. 情报学报, 2009,28(4):626－633.
[9] STEYVERS M, SMYTH P, ROSEN M, et al. Probabilistic author-topic models for information discovery[C]//Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. New York: ACM, 2004: 306－315.
[10] 李树青,孙颖. 一种基于 Web 请求访问模式和时间片震荡算法的便携式个性化服务框架[J]. 情报学报, 2014,33(3):228－238.
[11] 刘萍,郭月培,郭怡婷. 利用作者关键词网络探测作者相似性[J]. 现代图书情报技术, 2013,29(12):62－69.
[12] 巴志超,李纲,朱世伟. 基于语义网络的研究兴趣相似性度量方法[J]. 现代图书情报技术, 2016,32(4):81－90.
[13] 李纲,徐健,毛进,等. 合著作者研究兴趣相似性分布研究[J]. 图书情报工作, 2017,61(6):92－98.
[14] 关鹏,王曰芬. 学科领域生命周期中作者研究兴趣演化分析[J]. 图书情报工作, 2016,60(19):116－124.
[15] 丁学东. 文献计量学基础[M]. 北京:北京大学出版社,1992.
[16] 李纲,李岚凤,毛进,等. 作者合著网络中研究兴趣相似性实证研究[J]. 图书情报工作,2015,59(2):75－81.
[17] BLONDEL V D, GUILLAUME J L, LAMBIOTTE R, et al. Fast unfolding of communities in large networks[J]. Journal of statistical mechanics theory & experiment, 2008, 30(2):155－168.
[18] 唐磊,刘欢,文益民,等. 社会计算:社区发现和社会媒体挖掘[M]. 北京:机械工业出版社, 2012.

作者贡献说明:

徐健:论文撰写;
毛进:论文修改;
叶光辉:资料收集与整理;
巴志超:研究框架的设计与指导;
李纲:设计论文主题与思路。

Research on Community Membership Based on the Research Interest Similarity Network of Core Authors: Taking the Domestic Field of Information Science as an Example

Xu Jian[1] Mao Jin[1] Ye Guanghui[2] Ba Zhichao[1] Li Gang[1]

[1] Center for Studies of Information Resources, Wuhan University, Wuhan 430072

[2] School of Information Management of Central China Normal University, Wuhan 430079

Abstract: [**Purpose/significance**] This paper explores the community structure and community membership of the research interest similarity network of core authors in information science. [**Method/process**] We firstly download all papers of information science retrieved in CSSCI database using the China Library Category number. By recognizing the core -authors in this discipline with Price law, we compute the similarity between each two authors with bag-of-words model and construct the research interest similarity network of core authors. Then we divide it into four research community. Finally, we compute the every author's membership degree to different communities and his/her fuzzy entropy. [**Result/conclusion**] We discover that the domestic information science discipline has four research field: information organization and retrieval, bibliometrics and scientific evaluation, competitive intelligence and knowledge management, and the information science. Most authors' research isn't limited to one field. Finally, the authors in C2 and C3 merely take each other as secondary membership community, and it implies that the boundary between competitive intelligence and bibliometrics is very obvious.

Keywords: research interest network information science community membership degree similarity